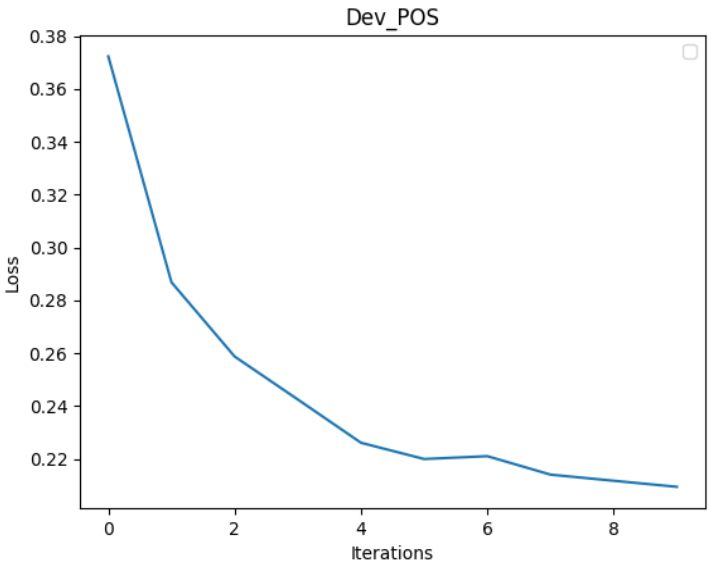
Deep Learning for Sequences – Assignment 2 – Part 1

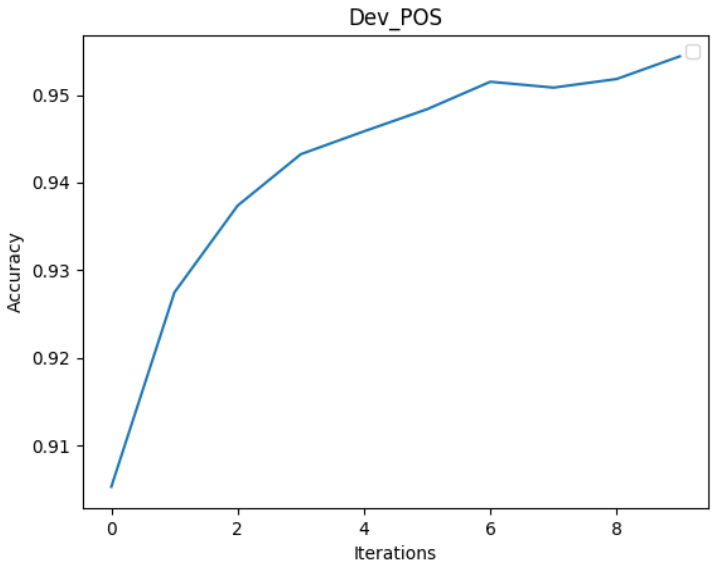
Daniel Juravski 206082323
Eyal Orbach 015369317

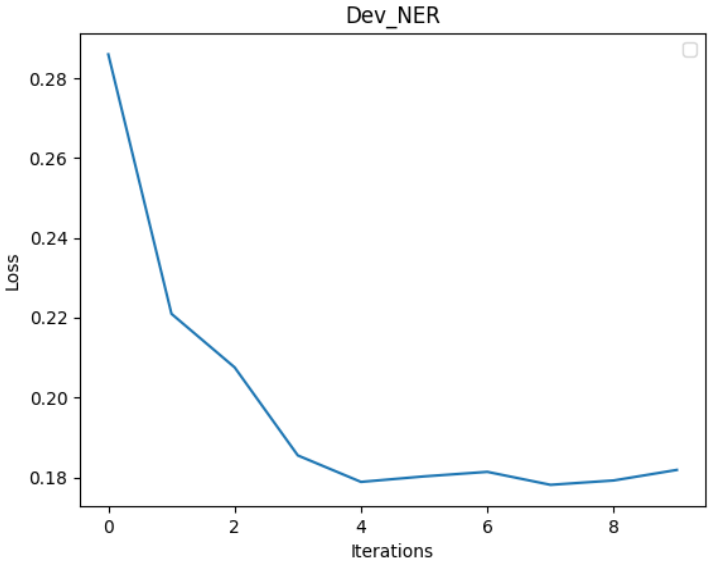
Our best parameters for both POS and NER:
Embedding vector size = 50
Ngram = 5
Hidden layer size = 50
Optimizer = “Adam”
Learning Rate = 0.002
Epochs = 5

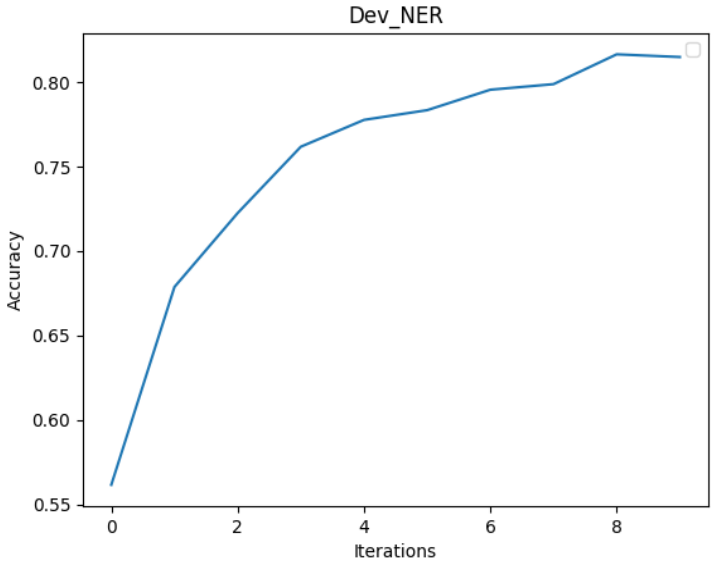
Considerations:

- There were words that appeared in the dev data but not in the train data, we figured it out with adding to the train data the 'UUUNKKK' word. Every word in the dev data that didn't have a representation in the embedding matrix, was represented as an 'UUUNKKK' word.
- In order to learn the behavior of unknown words, we decided to treat every word that appeared in the train data less than x times (x value was set to 2) as unknown word.
- To improve this we distinguished between four types of unknown words:
  - all capitals (UNK_ALLCAP)
  - the capital first letter (UNK_CapStart)
  - number (UNK_num)
  - all other unknown words (UUUNKKK)
- Each type received a vector and unknown words were mapped to their correct vector by checking these features.
- To deal with the vectors in places (-2,-1) and (+1,+2) for the first and last words in the sequence, we added 4 vectors to the E matrix [/S/S], [/S], [/E], [/E/E] that represented those places.









Notes:

1. Our model has learned in 5 epochs. We wanted to see the dev evaluation loss and accuracy some more times so we evaluate it twice in an epoch. Once in the middle of the epoch, and once at the end of the epoch.
2. The NER accuracy graph, calculated with the metric of ignoring the ‘O’ tag.